Complexity results of path following algorithms for linear programming which take into account the geometry of the central path

Renato Monteiro (Georgia Tech)
Takashi Tsuchiya (Inst. of Stat. Math.)

NUS, Singapore

Workshop on SDP and its Applications

January 11, 2005

Goals of the talk

- Understand the behavior of the central path and the Mizuno-Todd-Ye predictor-corrector (MTY P-C) algorithm for linear programming from the geometric point of view
- Estimate the iteration complexity of the MTY P-C algorithm in terms of the integral of a certain curvature of the central path
- Relate the above integral to a new iteration complexity bound for the MTY P-C algorithm involving a certain condition number of the constraint matrix A

Talk Outline

- **LP problem and assumptions;**
- **central path and its neighborhood;**
- **Mizuno-Todd-Ye predictor-corrector (MTY P-C) algorithm;**
- **condition number and scale-invariance;**
- **iteration complexity bounds for the MTY P-C alg.**
  - **classical one (1990)**
  - **new one (2003)**
- **illustrative LP instance**
- **curvature of the central path**
- **iteration complexity bounds in terms of a curvature integral**
- **directions for future research**

The LP problem

$$
\begin{array}{lll}
(\mathbf{P}) & \text{minimize}_{\mathbf{x}} & \mathbf{c}^T\mathbf{x} \\
& \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0},
\end{array}
$$

$$
\begin{array}{lll}
(D) & \text{maximize}_{(\mathbf{y},\mathbf{s})} & \mathbf{b}^T\mathbf{y} \\
& \text{subject to} & \mathbf{A}^T\mathbf{y} + \mathbf{s} = \mathbf{c},\ \mathbf{s} \geq \mathbf{0},
\end{array}
$$

Assumptions

1) (**P**) and (**D**) have interior-feasible solutions.
2) the rows of the $\mathbf{m} \times \mathbf{n}$ matrix $\mathbf{A}$ are linearly independent.

Definition: The duality gap of a feasible $\mathbf{w} = (\mathbf{x}, \mathbf{y}, \mathbf{s})$ is given by

$$\mathbf{c}^T\mathbf{x} - \mathbf{b}^T\mathbf{y} = (\mathbf{A}^T\mathbf{y} + \mathbf{s})^T\mathbf{x} - \mathbf{b}^T\mathbf{y} = \mathbf{x}^T\mathbf{s}.$$

Central path and its neighborhood

For each $\nu > \mathbf{0}$, the system

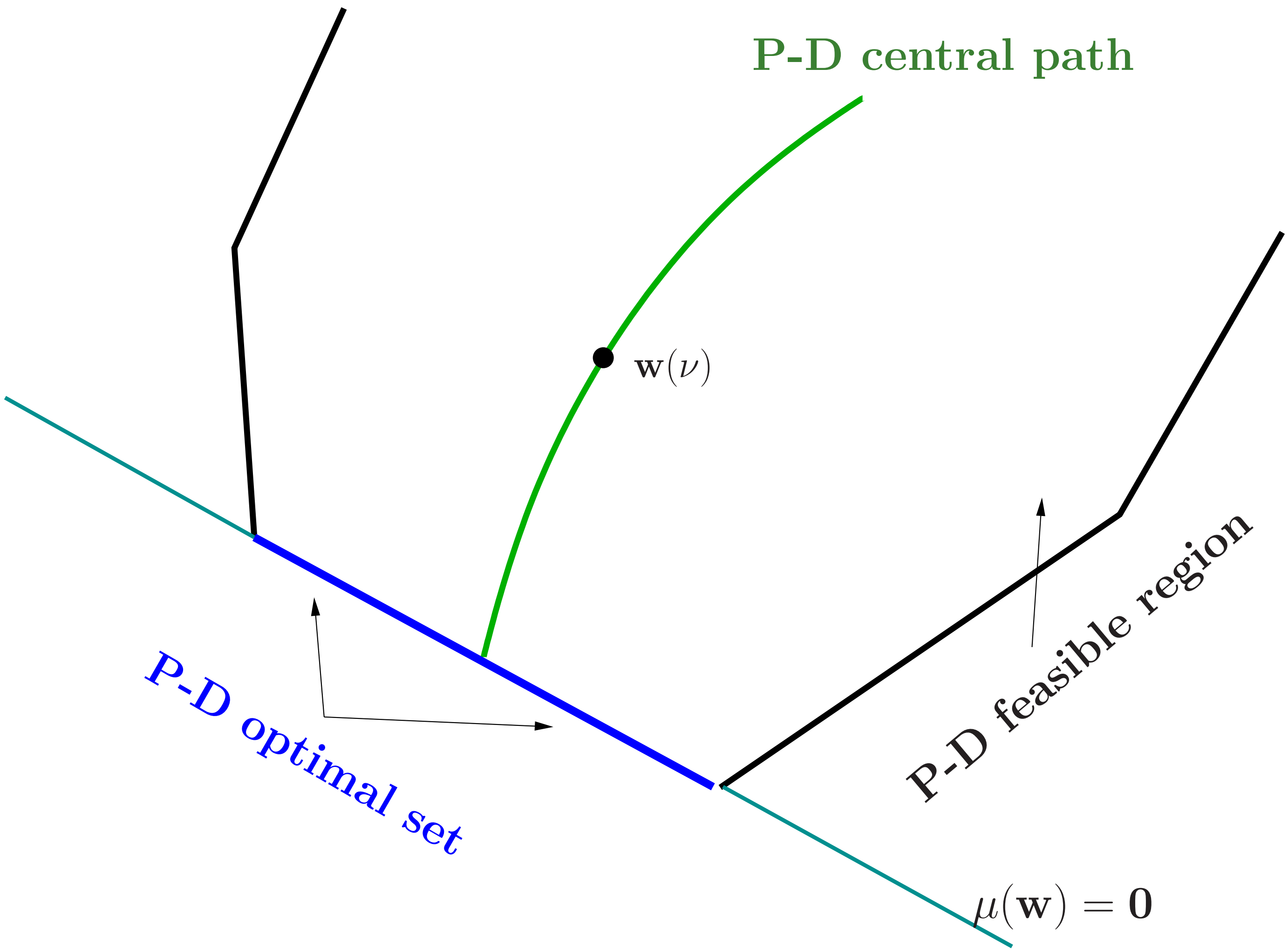
$$\begin{aligned} \mathbf{XS\,e} &= \nu\,\mathbf{e}, \\ \mathbf{Ax} - \mathbf{b} &= \mathbf{0}, \qquad (\mathbf{x}, \mathbf{s}) \geq \mathbf{0}, \\ \mathbf{A}^{\mathbf{T}}\mathbf{y} + \mathbf{s} - \mathbf{c} &= \mathbf{0}, \end{aligned}$$

where $\mathbf{X} = \mathbf{Diag}(\mathbf{x})$, $\mathbf{S} = \mathbf{Diag}(\mathbf{s})$ and $\mathbf{e} = (\mathbf{1}, \ldots, \mathbf{1})^{\mathbf{T}}$, has a unique solution $\mathbf{w}(\nu) = (\mathbf{x}(\nu), \mathbf{y}(\nu), \mathbf{s}(\nu))$, which converges to a primal-dual optimal solution as $\nu \to \mathbf{0}$.

The MTY P-C is based on the 2-norm neighborhood of the central path:

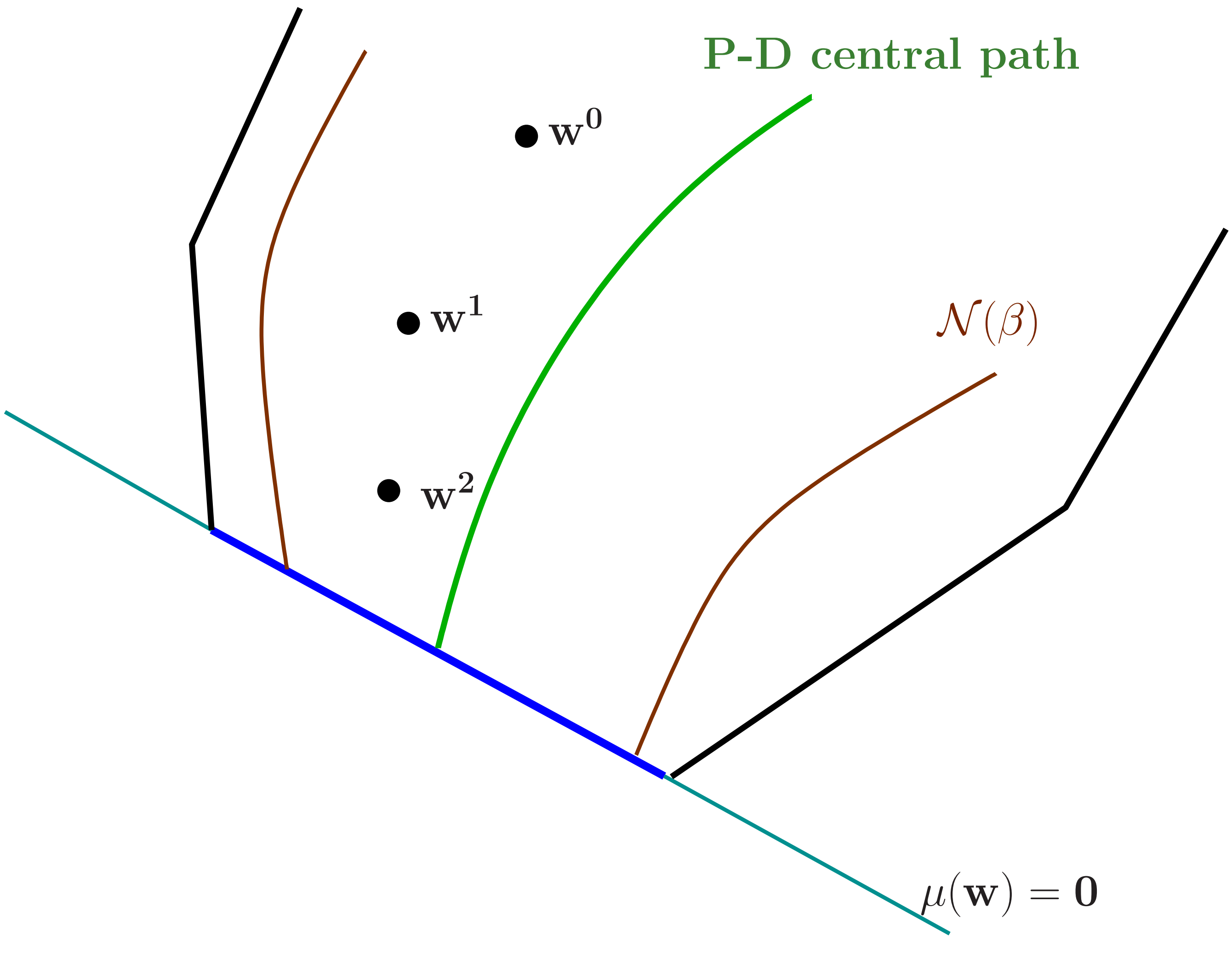
$$\mathcal{N}(\beta) \equiv \{\mathbf{w} = (\mathbf{x}, \mathbf{y}, \mathbf{s}) \textbf{ feasible} : \|\mathbf{Xs} - \mu\mathbf{e}\| \leq \beta\mu\},$$

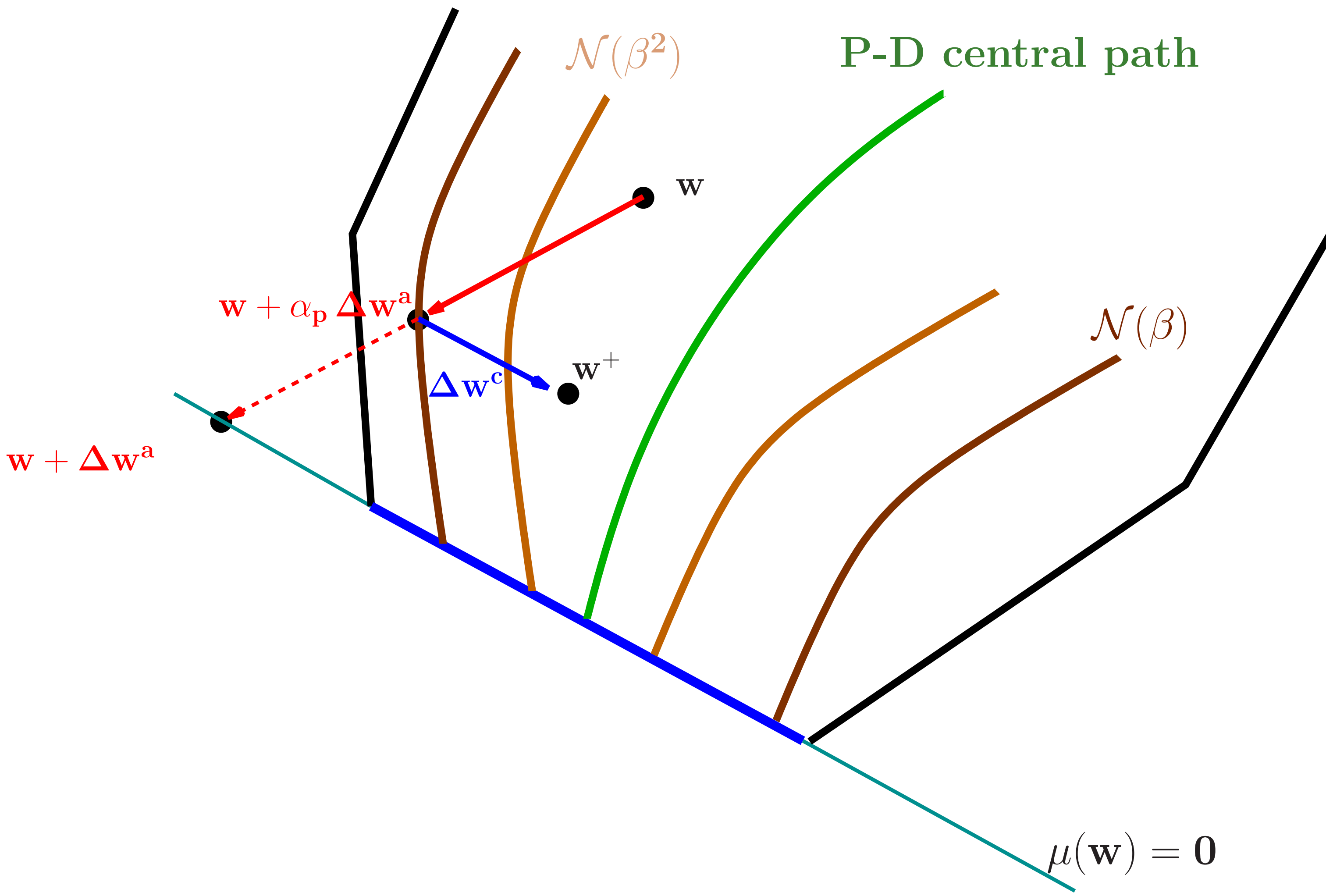
where $\mu = \mu(\mathbf{w}) \equiv (\mathbf{x}^{\mathbf{T}}\mathbf{s})/\mathbf{n}$ and $\beta \in (\mathbf{0}, \mathbf{1})$ is a fixed constant.



$$\mathbf{w}(\nu) \quad = \quad (\mathbf{x}(\nu), \mathbf{s}(\nu), \mathbf{y}(\nu))$$

$$\mu(\mathbf{w}) \quad := \quad \frac{\mathbf{c}^{\mathbf{T}}\mathbf{x} - \mathbf{b}^{\mathbf{T}}\mathbf{y}}{\mathbf{n}} \;=\; \frac{\mathbf{s}^{\mathbf{T}}\mathbf{x}}{\mathbf{n}}$$


P-D central path
$\mathbf{w}^0$
$\mathbf{w}^1$
$\mathbf{w}^2$
$\mathcal{N}(\beta)$
$\mu(\mathbf{w}) = \mathbf{0}$



$\mathcal{N}(\beta^2)$
P-D central path
$\mathbf{w}$
$\mathbf{w} + \alpha_p \, \boldsymbol{\Delta} \mathbf{w}^a$
$\mathcal{N}(\beta)$
$\mathbf{w}^+$
$\boldsymbol{\Delta} \mathbf{w}^c$
$\mathbf{w} + \boldsymbol{\Delta} \mathbf{w}^a$
$\mu(\mathbf{w}) = \mathbf{0}$

Search Directions

For a strictly feasible $\mathbf{w} = (\mathbf{x}, \mathbf{y}, \mathbf{s})$, the Newton direction $\mathbf{\Delta w} = (\mathbf{\Delta x}, \mathbf{\Delta y}, \mathbf{\Delta s})$ towards the point $\mathbf{w}(\nu) = (\mathbf{x}(\nu), \mathbf{y}(\nu), \mathbf{s}(\nu))$ is the solution of

$$\begin{aligned} \mathbf{X\Delta s} + \mathbf{S\Delta x} &= -\mathbf{Xs} + \nu\mathbf{e} \\ \mathbf{A\Delta x} &= \mathbf{0} \\ \mathbf{A^T\Delta y} + \mathbf{\Delta s} &= \mathbf{0} \end{aligned}$$

Setting $\nu = \mathbf{0}$ yields the predictor (or affine scaling) direction at $\mathbf{w}$.

Setting $\nu = \mu(\mathbf{w})$ yields the corrector (or centrality) direction at $\mathbf{w}$.

An iteration of the MTY P-C Alg.

Let $\mathbf{w} = (\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{N}(\beta^2)$ be given, where $\beta \in (0, 1/2]$.

1) **Compute the AS direction $\Delta\mathbf{w^a} = (\Delta\mathbf{x^a}, \Delta\mathbf{y^a}, \Delta\mathbf{s^a})$ at $\mathbf{w}$;**

2) **Let $\alpha_\mathbf{p} > 0$ be the largest $\alpha \in [0, 1]$ such that $\mathbf{w} + \alpha\Delta\mathbf{w^a} \in \mathcal{N}(\beta)$;**

3) **Set $\mathbf{w_p} = \mathbf{w} + \alpha_\mathbf{p}\Delta\mathbf{w^a}$;**

4) **Compute the corrector direction $\Delta\mathbf{w^c} = (\Delta\mathbf{x^c}, \Delta\mathbf{y^c}, \Delta\mathbf{s^c})$ at $\mathbf{w_p}$;**

5) **The next point $\mathbf{w}^+$ is determined as $\mathbf{w}^+ = \mathbf{w_p} + \Delta\mathbf{w^c}$;**

It can be proved that $\mathbf{w}^+ \in \mathcal{N}(\beta^2)$. Hence, a new iteration can be started by setting $\mathbf{w} \leftarrow \mathbf{w}^+$ and going back to 1).

THE CONDITION NUMBER $\bar{\chi}_{\mathbf{A}}$

Define

$$\bar{\chi}_{\mathbf{A}} \equiv \sup\{\|(\mathbf{A}\mathbf{D}\mathbf{A}^{\mathbf{T}})^{-\mathbf{1}}\mathbf{A}\mathbf{D}\| : \mathbf{D} \in \mathcal{D}\},$$

where $\mathcal{D}$ denotes the set of all positive definite diagonal matrices.

Facts:

1) $\bar{\chi}_{\mathbf{A}} = \max\{\|\mathbf{B}^{-\mathbf{1}}\mathbf{A}\| : \mathbf{B}$ **is a basis of A**$\}$.

2) **Finding an upper bound for $\bar{\chi}_{\mathbf{A}}$ is a $\mathcal{NP}$ hard problem.**

3) **If A integral then $\bar{\chi}_{\mathbf{A}} \leq \mathbf{2}^{\mathbf{L}_{\mathbf{A}}}$, where $\mathbf{L}_{\mathbf{A}}$ is the input size of A.**

Scale Invariance

Let $\mathbf{D}$ be a positive diagonal matrix and consider the pair of LPs:

$$
\begin{array}{lll}
(\tilde{\mathbf{P}}) & \text{minimize} & (\mathbf{Dc})^{\mathbf{T}}\tilde{\mathbf{x}} \\
 & \text{subject to} & \mathbf{AD}\tilde{\mathbf{x}} = \mathbf{b},\ \tilde{\mathbf{x}} \geq \mathbf{0}, \\
(\tilde{\mathbf{D}}) & \text{maximize} & \mathbf{b}^{\mathbf{T}}\tilde{\mathbf{y}} \\
 & \text{subject to} & \mathbf{DA}^{\mathbf{T}}\tilde{\mathbf{y}} + \tilde{\mathbf{s}} = \tilde{\mathbf{c}},\ \tilde{\mathbf{s}} \geq \mathbf{0},
\end{array}
$$

obtained from $(\mathbf{P})$ and $(\mathbf{D})$ by performing the change of variables $(\mathbf{x}, \mathbf{y}, \mathbf{s}) = \mathbf{\Phi}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{s}}) \equiv (\mathbf{D}\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \mathbf{D}^{-1}\tilde{\mathbf{s}})$.

The MTY P-C algorithm is scaling-invariant, i.e., if $\{\mathbf{w}^{\mathbf{k}}\}$ and $\{\tilde{\mathbf{w}}^{\mathbf{k}}\}$ denote the sequence of iterates generated by the MTY P-C algorithm in the original and the scaled space, then $\mathbf{w}^{\mathbf{k}} = \mathbf{\Phi}(\tilde{\mathbf{w}}^{\mathbf{k}})$ for all $\mathbf{k} \geq \mathbf{1}$, as long as $\mathbf{w}^{\mathbf{0}} = \mathbf{\Phi}(\tilde{\mathbf{w}}^{\mathbf{0}})$.

ITERATION-COMPLEXITY BOUNDS

Given $\mathbf{0} < \nu_\mathbf{f} < \nu_\mathbf{i}$, denote by $\mathbf{N}(\nu_\mathbf{i}, \nu_\mathbf{f}, \beta)$ the largest possible number of iterations required by the MTY P-C algorithm to find an iterate with duality gap $\leq \nu_\mathbf{f}$ when started from any $\mathbf{w^0} \in \mathcal{N}(\beta^\mathbf{2})$ such that $\mu(\mathbf{w^0}) = \nu_\mathbf{i}$.

Classical Result: For any $\beta \in (\mathbf{0}, \mathbf{1}/\mathbf{2}]$,

$$\sqrt{\beta} \cdot \mathbf{N}(\nu_\mathbf{i}, \nu_\mathbf{f}, \beta) \leq \sqrt{\mathbf{n}} \log \left(\frac{\nu_\mathbf{i}}{\nu_\mathbf{f}} \right)$$

Lemma: Suppose $\mathbf{w} \in \mathcal{N}(\beta^\mathbf{2})$, where $\beta \in (\mathbf{0}, \mathbf{1}/\mathbf{2}]$. Then, $\mathbf{w}^+ \in \mathcal{N}(\beta^\mathbf{2})$ and

$$\frac{\mu(\mathbf{w}^+)}{\mu(\mathbf{w})} \leq \mathbf{1} - \sqrt{\frac{\beta}{\mathbf{n}}}$$

Vavasis-Ye Algorithm

Iteration Complexity Bound: The number of iterations to solve a linear program is

$$\mathcal{O}(\mathbf{n}^{\mathbf{3.5}} \log(\mathbf{n} + \bar{\chi}_{\mathbf{A}}))$$

Note: Their bound does not depend on $\nu_{\mathbf{i}}$ and $\nu_{\mathbf{f}}$!

Their algorithm accelerates an ordinary primal-dual path following method (e.g., the MTY P-C algorithm) by using from time to time a step called the layered-least-square step.

V-Y algorithm is not scaling invariant.

New complexity for the MTY method

Theorem (Monteiro and Tsuchiya 2003): For any $\beta \in (\mathbf{0}, \mathbf{1}/\mathbf{2}]$,

$$\mathbf{N}(\nu_{\mathbf{i}}, \nu_{\mathbf{f}}, \beta) = \mathcal{O}\left(\mathbf{T}(\nu_{\mathbf{i}}/\nu_{\mathbf{f}}) + \mathbf{n}^{\mathbf{3.5}} \log(\bar{\chi}^*_{\mathbf{A}} + \mathbf{n})\right)$$

iterations, where $\bar{\chi}^*_{\mathbf{A}} \equiv \inf\{\bar{\chi}_{\mathbf{AD}} : \mathbf{D} \in \mathcal{D}\}$ and

$$\mathbf{T}(\eta) \equiv \min\left\{\mathbf{n}^{\mathbf{2}} \log(\log \eta), \log \eta\right\}$$

Remark: In contrast to $\bar{\chi}_{\mathbf{A}}$, the quantity $\bar{\chi}^*_{\mathbf{A}}$ is scaling invariant. Usually $\bar{\chi}^*_{\mathbf{A}} << \bar{\chi}_{\mathbf{A}}$. Hence, the above complexity is not comparable to the one associated with the V-Y method.

Lemma: For any $\beta \in (\mathbf{0}, \mathbf{1}/\mathbf{2}]$ and $\mathbf{w} \in \mathcal{N}(\beta^{\mathbf{2}})$:

$$\frac{\mu(\mathbf{w}^+)}{\mu(\mathbf{w})} \leq \frac{\kappa(\mathbf{w})^{\mathbf{2}}}{\beta},$$

where

$$\kappa(\mathbf{w}) := \left(\frac{\|\mathbf{\Delta x^a}(\mathbf{w})\mathbf{\Delta s^a}(\mathbf{w})\|}{\mu(\mathbf{w})}\right)^{\mathbf{1}/\mathbf{2}}$$

Consequences

Under the Turing machine model, the iteration-complexity of the MTY P-C algorithm is

$$\begin{aligned} & \mathcal{O}(\mathbf{n^{3.5}L_A} + \min\{\mathbf{L}, \mathbf{n^2} \log \mathbf{L}\}) \\ \leq \; & \mathcal{O}(\mathbf{n^{3.5}L_A} + \mathbf{L}) \end{aligned}$$

Given $\mathbf{A}$, there exist many nontrivial $(\mathbf{b}, \mathbf{c})$ for which the complexity of the MTY P-C algorithm for solving $(\mathbf{P})$ and $(\mathbf{D})$ is $\mathcal{O}(\mathbf{L})$

Example

Consider the LP

$$\max\{\mathbf{b}^{\mathbf{T}}\mathbf{y} : \mathbf{A}^{\mathbf{T}}\mathbf{y} \le \mathbf{c}\},$$

where

$$\mathbf{A} = \begin{pmatrix} 0 & \frac{\sqrt{6}}{3} & -\frac{\sqrt{6}}{3} & \mathbf{0} \\ 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{3} & -\frac{2\sqrt{2}}{3} \\ -1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix},$$

$$\mathbf{b} = \begin{pmatrix} -10^{-9} \\ -10^{-5} \\ -1 \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} 0 \\ \frac{2\sqrt{6}}{3} \\ 0 \\ 0 \end{pmatrix}.$$

Example (continued)

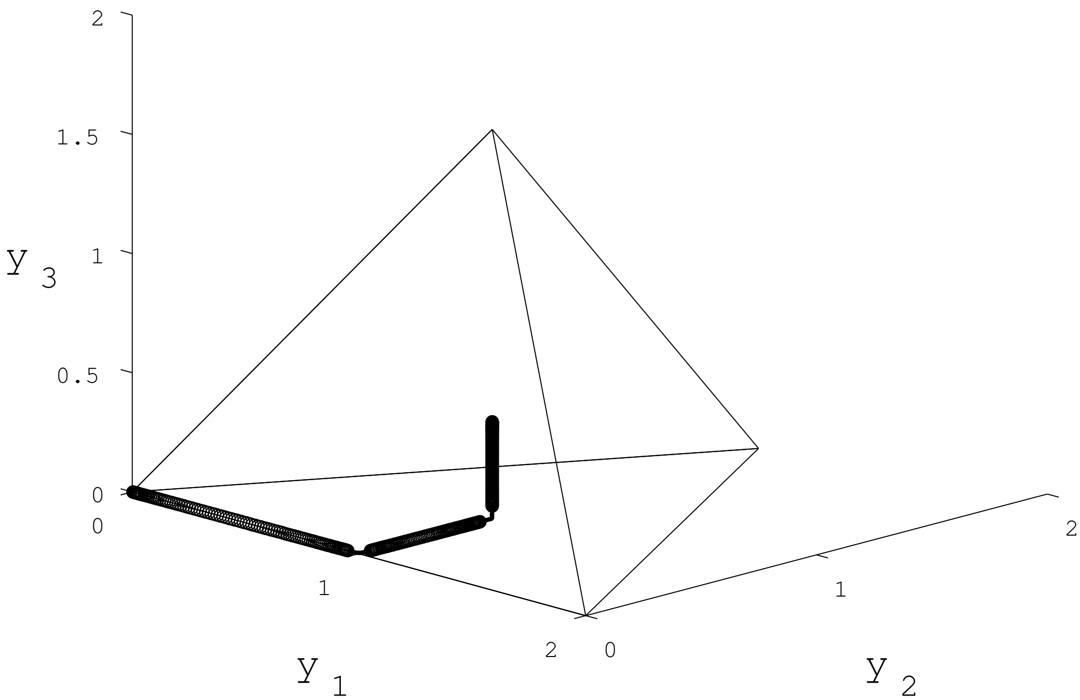


Figure 1: Figure for the LP instance

Example (continued)

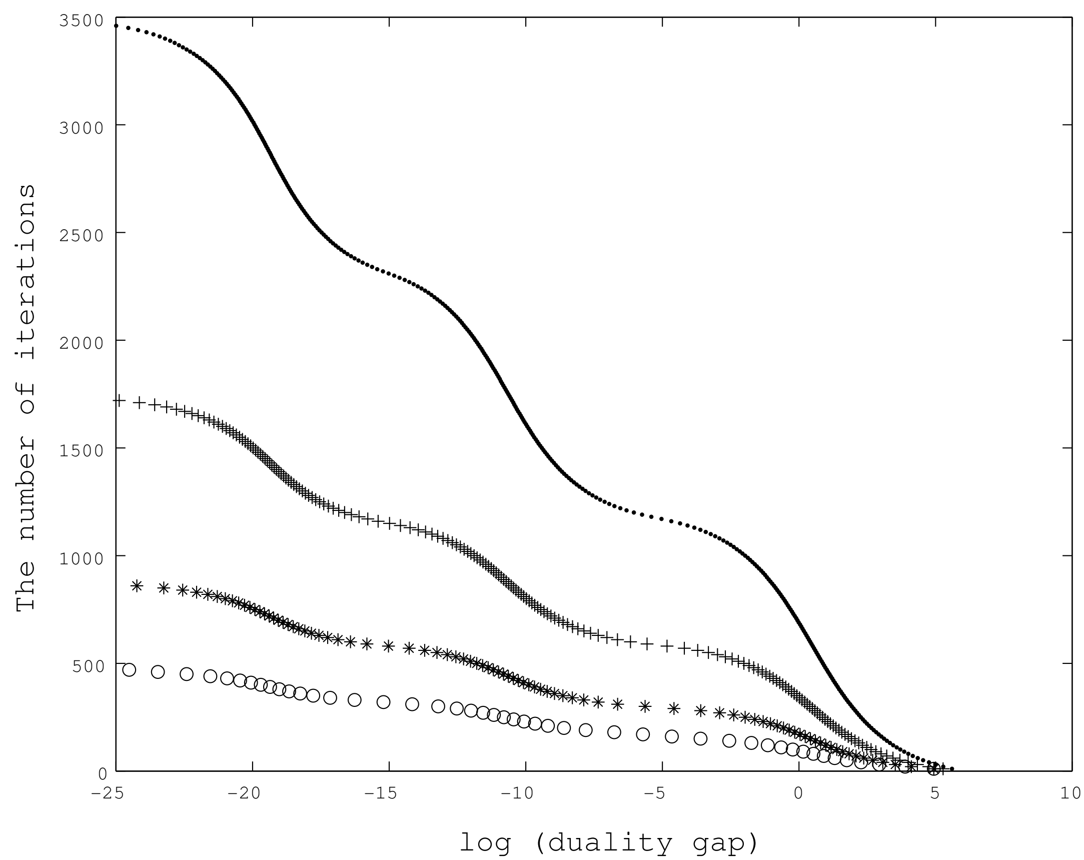


Figure 2: $\log \mu$ **versus** $\mathbf{N}(\nu_{\mathbf{i}}, \mu, \beta)$ **(**$\cdot : \sqrt{\beta} =$ **0.0025**; $+ : \sqrt{\beta} =$ **0.005**; $* : \sqrt{\beta} =$ **0.01**; $\circ : \sqrt{\beta} =$ **0.02)**

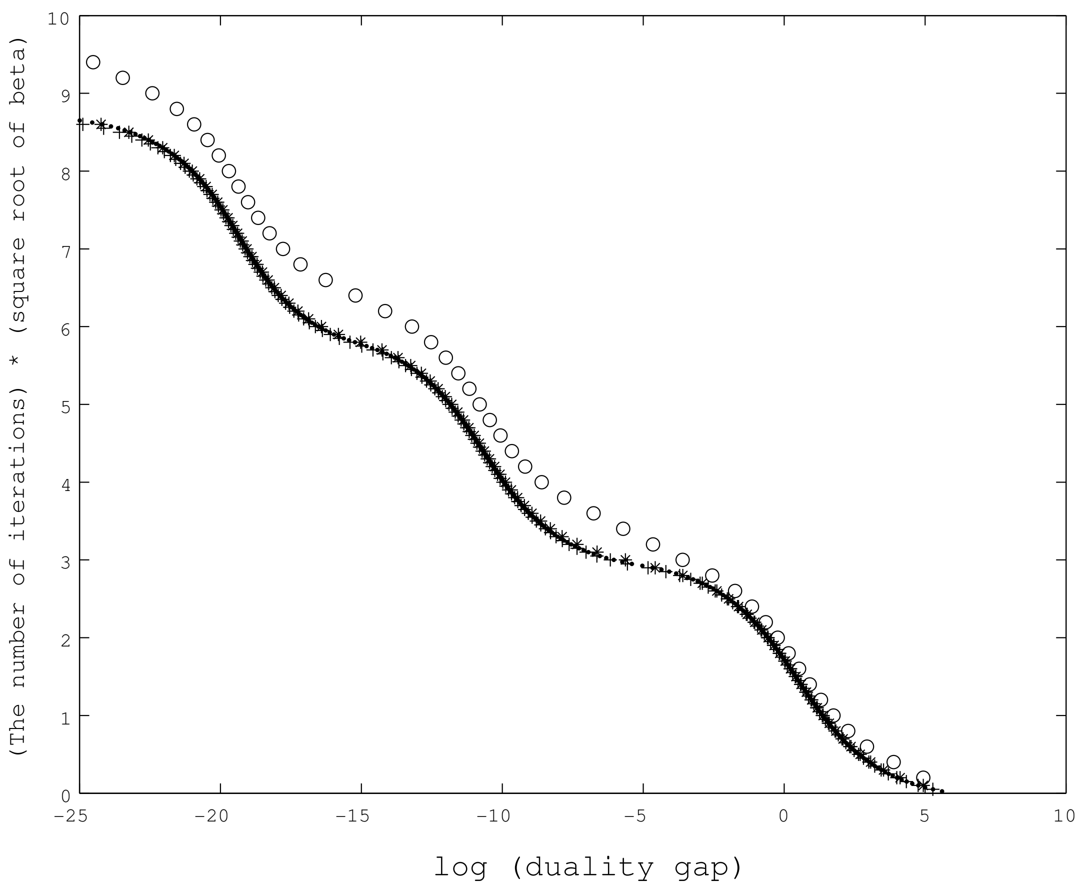


Figure 3: $\log \mu$ **versus** $\sqrt{\beta} \cdot \mathbf{N}(\nu_{\mathbf{i}}, \mu, \beta)$ **(**$\cdot : \sqrt{\beta} =$ **0.0025**; $+ : \sqrt{\beta} =$ **0.005**; $* : \sqrt{\beta} =$ **0.01**; $\circ : \sqrt{\beta} =$ **0.02)**

Question: Does $\sqrt{\beta} \cdot \mathbf{N}(\nu_{\mathbf{i}}, \mu, \beta)$ **always converge as** $\beta \to 0$**?**

Example (continued)

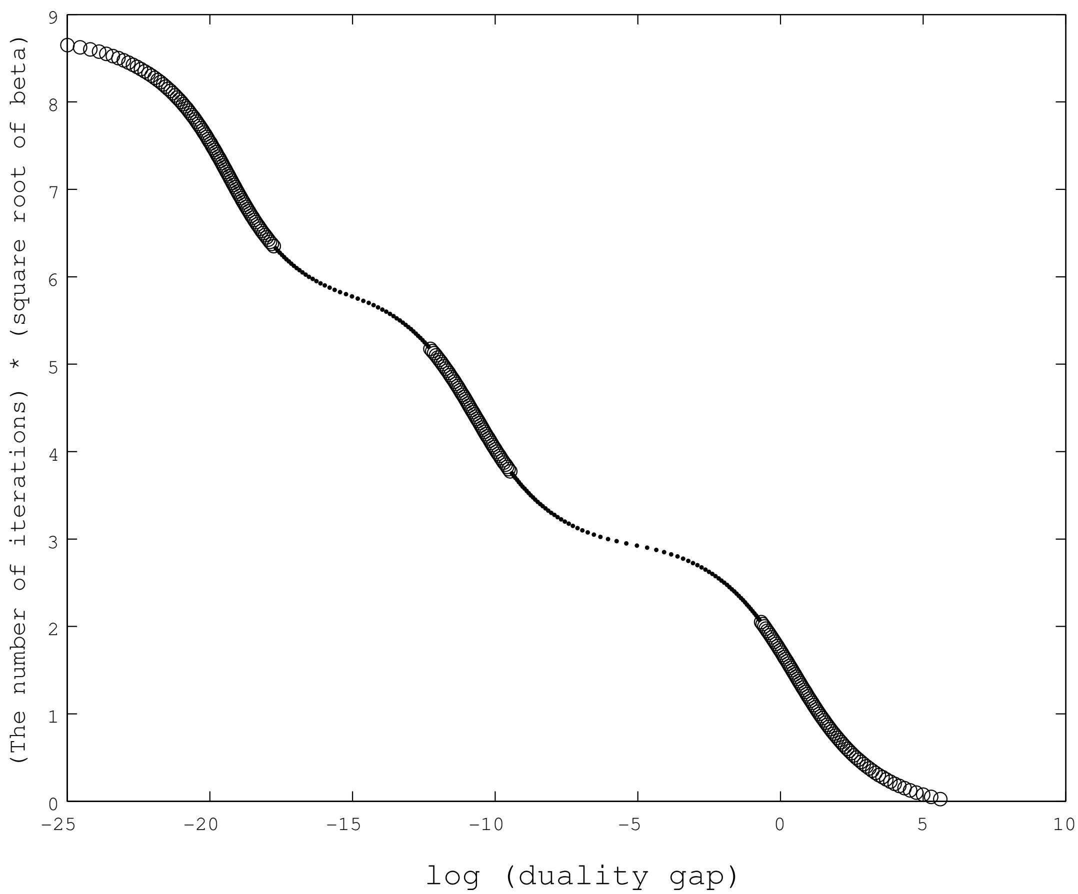


Figure 4: $\log \mu$ **versus** $\sqrt{\beta} \cdot \mathbf{N}(\nu_{\mathbf{i}}, \mu, \beta)$ **(The big dots correspond to the ones in Figure 1.)**

Question: How to define straight and curved parts of the central path?

Curvature of the central path

Definition: **The curvature of the central path is the function** $\kappa : (\mathbf{0}, \infty) \to [\mathbf{0}, \infty)$ **defined as**

$$\kappa(\nu) \equiv \|\nu\dot{\mathbf{x}}(\nu)\dot{\mathbf{s}}(\nu)\|^{\mathbf{1/2}}, \quad \forall \nu > \mathbf{0}.$$

Note: **if** $\mathbf{w} = \mathbf{w}(\nu)$ **then** $\kappa(\mathbf{w}) = \kappa(\nu)$

For a given $\nu > \mathbf{0}$ **and** $\beta \in (\mathbf{0}, \mathbf{1})$, **define**

$$\mathcal{T}(\beta, \nu) \equiv \{\mathbf{t} \in \Re : \mathbf{w}(\nu) - \mathbf{t}\nu\dot{\mathbf{w}}(\nu) \in \mathcal{N}(\beta)\}$$

Note that $\mathbf{w}(\nu) - \mathbf{t}\nu\dot{\mathbf{w}}(\nu) \approx \mathbf{w}((\mathbf{1} - \mathbf{t})\nu)$.

Proposition: $\mathcal{T}(\beta, \nu)$ **is a closed interval and**

$$\lim_{\beta \downarrow \mathbf{0}} \frac{\textbf{length of } \mathcal{T}(\beta, \nu)}{\sqrt{\beta}} = \frac{\mathbf{2}}{\kappa(\nu)}$$

Complexity in terms of the curvature

Theorem (Sonnevend, Stoer and Zhao 1994):

$$\mathbf{N}(\nu_{\mathbf{i}}, \nu_{\mathbf{f}}, \beta) = \mathcal{O}\left(\int_{\nu_{\mathbf{f}}}^{\nu_{\mathbf{i}}} \frac{\kappa(\nu)}{\nu} \mathbf{d}\nu + \log\left(\frac{\nu_{\mathbf{i}}}{\nu_{\mathbf{f}}}\right)\right).$$

Note: **Since $\kappa(\nu) \leq \sqrt{\mathbf{n}/\mathbf{2}}$ for all $\nu > \mathbf{0}$, the classical bound follows from the above bound.**

Theorem 1 (Monteiro and Tsuchiya 2005):

$$\begin{aligned} \lim_{\beta \to \mathbf{0}} \sqrt{\beta} \cdot \mathbf{N}(\nu_{\mathbf{i}}, \nu_{\mathbf{f}}, \beta) &= \int_{\nu_{\mathbf{f}}}^{\nu_{\mathbf{i}}} \frac{\kappa(\nu)}{\nu} \mathbf{d}\nu \\ &\leq \sqrt{\mathbf{n}} \log\left(\frac{\nu_{\mathbf{i}}}{\nu_{\mathbf{f}}}\right) \end{aligned}$$

Recall that one of the M-T bounds is

$$\mathbf{N}(\nu_{\mathbf{i}}, \nu_{\mathbf{f}}, \beta) = \mathcal{O}\left(\mathbf{n}^{\mathbf{3.5}} \log(\bar{\chi}_{\mathbf{A}}^* + \mathbf{n}) + \log\left(\frac{\nu_{\mathbf{i}}}{\nu_{\mathbf{f}}}\right)\right).$$

Bound on the curvature integral

Theorem 2 (Monteiro and Tsuchiya 2005):

For every $0 < \nu_f < \nu_i$, we have:

$$\int_{\nu_f}^{\nu_i} \frac{\kappa(\nu)}{\nu} \mathbf{d}\nu \le \mathcal{O}\left(\mathbf{n}^{3.5} \log(\bar{\chi}^*_{\mathbf{A}} + \mathbf{n})\right)$$

Hence,

$$\int_{\mathbf{0}}^{\infty} \frac{\kappa(\nu)}{\nu} \mathbf{d}\nu \le \mathcal{O}\left(\mathbf{n}^{3.5} \log(\bar{\chi}^*_{\mathbf{A}} + \mathbf{n})\right)$$

Geometry of the central path

Vavasis and Ye 1996: **"The central path consists of $\mathcal{O}(n^2)$ long and straight parts and other curved parts"**

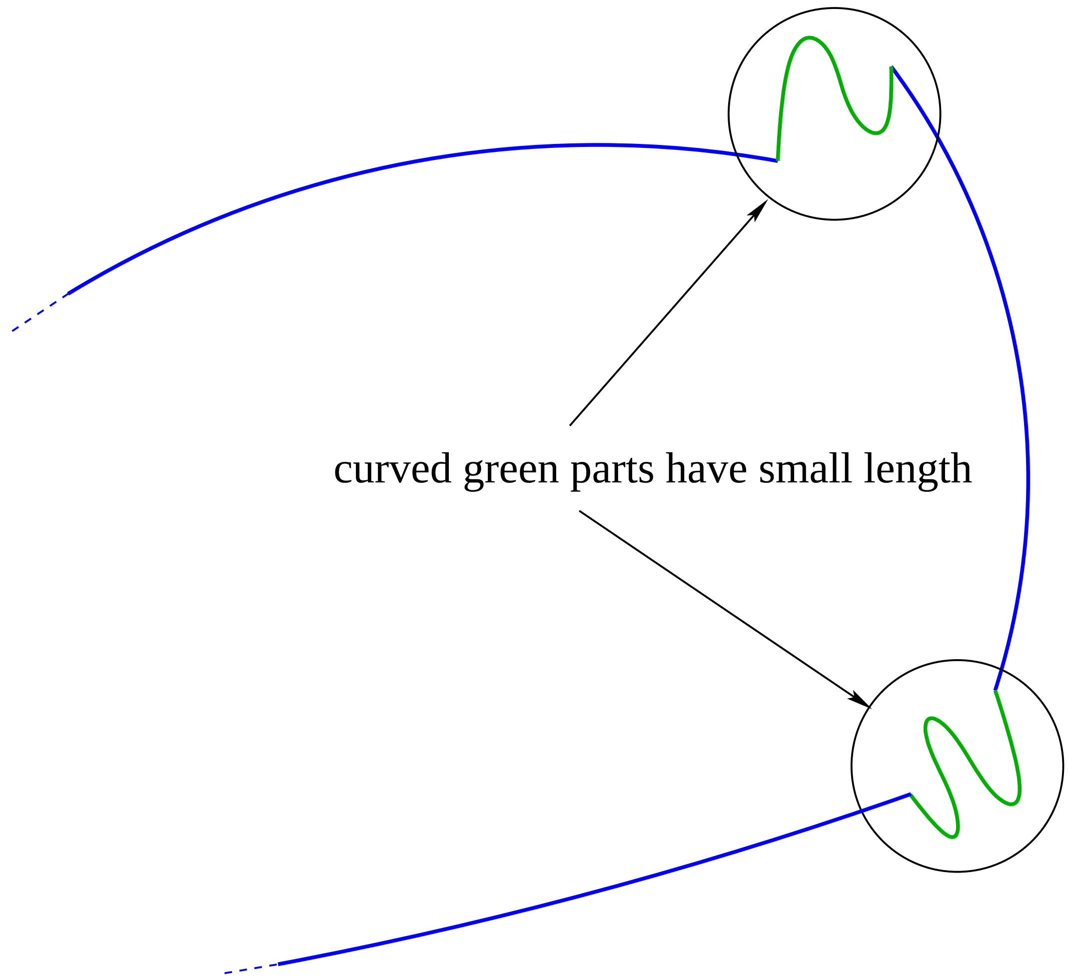
We want to formally establish this statement!

Theorem 3: **For any $\bar{\kappa} \in (\mathbf{0}, \sqrt{\mathbf{n}/\mathbf{2}})$, there exist $\mathbf{l} \leq \mathbf{n}(\mathbf{n}-\mathbf{1})/\mathbf{2}$ closed intervals $\mathbf{I_k}$ such that:**

a) $\{\nu > \mathbf{0} : \kappa(\nu) \geq \bar{\kappa}\} \subseteq \cup_{\mathbf{k}=\mathbf{1}}^{\mathbf{l}} \mathbf{I_k}$
(union of $\mathbf{I_k}$'s covers portion with large curvature)

b) **the logarithmic length of each $\mathbf{I_k}$ is bounded by** $\mathcal{O}\left(\mathbf{n} \log(\bar{\chi}_{\mathbf{A}}^* + \mathbf{n}) + \mathbf{n} \log \bar{\kappa}^{-\mathbf{1}}\right)$
(independent of b and c)

Geometric illustration of the C-P



The blue parts are long but quite straight!

The MTY P-C algorithm converges R-quadratically over the blue parts.

There are at most $\mathcal{O}(n^2)$ blue and green parts.

Directions for future research

- **Generalizations to other cone programming problems such as SOCP and SDP**
- **Are infeasible path following methods ammenable to the same kind of analysis? Can new iteration complexity bounds be obtained for them?**
- **Is it possible to interpret the curvature $\kappa(\nu)$ as the one used in differential geometry? What further insights can be gained through this approach?**
- **Can an iteration complexity bound depending only on $\mathbf{n}$ and $\bar{\chi}^*_{\mathbf{A}}$ be derived for the MTY P-C algorithm?**
- **Is it possible to derive a Zhao and Stoer's type result with $\log\log$, i.e.**

$$\mathbf{N}(\nu_{\mathbf{i}}, \nu_{\mathbf{f}}, \beta) = \mathcal{O}\left(\int_{\nu_{\mathbf{f}}}^{\nu_{\mathbf{i}}} \frac{\kappa(\nu)}{\nu} \mathbf{d}\nu + \mathbf{n^2} \log\log\left(\frac{\nu_{\mathbf{i}}}{\nu_{\mathbf{f}}}\right)\right).$$